

# Clasificación de servicios Web mediante una red neuronal artificial usando $n$ -gramas de palabras

José A. Reyes-Ortiz, Maricela Bravo

Universidad Autónoma Metropolitana, Unidad Azcapotzalco,  
Departamento de Sistemas, Distrito Federal,  
México

{jaro, mcbc}@correo.azc.uam.mx

**Resumen.** Este artículo presenta un enfoque basado en  $n$ -gramas de palabras para la clasificación automática de servicios Web utilizando una red neuronal artificial de tipo perceptrón multicapa. Los servicios Web contienen información de gran utilidad para lograr una clasificación basada en la funcionalidad del mismo. El enfoque se basa en  $n$ -gramas de palabras extraídas de la descripción del servicio web para determinar su pertenencia a una categoría. La experimentación realizada muestra resultados prometedores, logrando una clasificación con una medida  $F=0.91$  utilizando unigramas ( $1$ -gramas) de palabras (características compuestas por una unidad léxica) y una ponderación  $TF-IDF$ .

**Palabras clave:** Servicios web, modelo espacio vectorial, clasificación de servicio web,  $n$ -gramas de palabras, red neuronal perceptrón multicapa.

## 1. Introducción

Los servicios Web son componentes de software reutilizables a través de los cuales es posible construir e integrar nuevas aplicaciones sin tener que implementar todos los elementos de un sistema. Actualmente los servicios Web han cobrado mayor popularidad debido a su proliferación para el ofrecimiento de servicios de almacenamiento y administración de recursos en la nube. Los servicios Web se encuentran disponibles tanto en repositorios públicos como privados a través de las descripciones de los servicios. Existen varios repositorios de servicios Web públicos: a) el directorio de Servicios Web SOAP soportado por Membrane<sup>1</sup>; b) el repositorio de servicios Web Visual Web Service<sup>2</sup>; c) el repositorio de servicios Web XMethods<sup>3</sup>; d) ProgrammableWeb<sup>4</sup>; e) OWLS-TC<sup>5</sup> es una colección de servicios de prueba

---

<sup>1</sup> <http://www.service-repository.com/>

<sup>2</sup> <http://www.visualwebservice.com>

<sup>3</sup> <http://www.xmethods.com>

<sup>4</sup> <http://www.programmableweb.com>

<sup>5</sup> <http://projects.semwebcentral.org/projects/owls-tc/>

recuperados con sus respectivas anotaciones en OWL-S. Estas anotaciones expresan la semántica de los elementos de un servicio Web.

Las descripciones de servicios Web se realizan mediante el lenguaje estándar WSDL, dicha descripción consiste de un archivo de texto basado en XML, dentro del cual se definen los elementos necesarios para la invocación del servicio.

Para que los programadores y desarrolladores de aplicaciones puedan hacer uso de los servicios necesitan buscarlos dentro de los repositorios de servicios. Esta tarea se le conoce comúnmente como descubrimiento de servicios. Sin embargo, el descubrimiento de servicios sigue siendo una tarea ardua y propensa a errores, ya que la mayoría de los repositorios ofrecen mecanismos de búsqueda basados en coincidencia de palabras clave. Aunado a este problema está el hecho de que los repositorios de servicios están organizados principalmente mediante estructuras estáticas que no permiten una organización flexible y dinámica de los servicios.

Este trabajo de investigación tiene como objetivo principal mejorar la estructura organizacional de los repositorios de servicios Web de tal forma que se facilite el descubrimiento de servicios. La principal contribución de este artículo se centra en un algoritmo de clasificación de servicios Web utilizando *n-gramas* de palabras. Como resultado se obtienen colecciones de servicios Web organizados por temas, su búsqueda se agiliza, consumiendo menos recursos, debido a que se realiza entre servicios dentro de la misma categoría.

El resto del artículo está organizado de la siguiente manera. En la Sección 2 se presenta la descripción de servicio Web. La Sección 3 expone un acercamiento al estado de arte de los trabajos relacionados. La Sección 4 muestra el enfoque propuesto. La Sección 5 presenta la experimentación y los resultados obtenidos para la clasificación de servicios web. Finalmente, las conclusiones son presentadas en la Sección 6.

## 2. Descripción de servicios web

El lenguaje de descripción de servicios recomendado para la implementación de servicios Web es llamado *Web Service Description Language (WSDL)*, el cual es, actualmente, un estándar aceptado por la *W3C*. *WSDL* define una gramática XML para describir servicios conectados como una colección de nodos de comunicación capaces de intercambiar mensajes<sup>6</sup>.

En este trabajo son consideradas la versión *WSDL 2.0*, la más reciente y que incorpora cambios significados en la descripción del servicio con respecto a la versión *WSDL 1.1*. *WSDL 2.0* cambia la etiqueta *<definitions>* por la etiqueta *<description>* (ver Fig. 1). Las diferencias más significativas entre *WSDL 1.1* y *WSDL 2.0* son: el *targetNameSpace* es un atributo requerido en *WSDL 2.0*; la construcción de mensajes es eliminada en *WSDL 2.0*; la sobrecarga de operadores no es soportada en *WSDL 2.0*; *PortType* ha cambiado de nombre a *Interface*; la herencia en *Interface* es soportada mediante el uso de atributos extendidos; y *Port* ha cambiado el nombre por *Endpoint*.

---

<sup>6</sup> <http://www.w3.org>

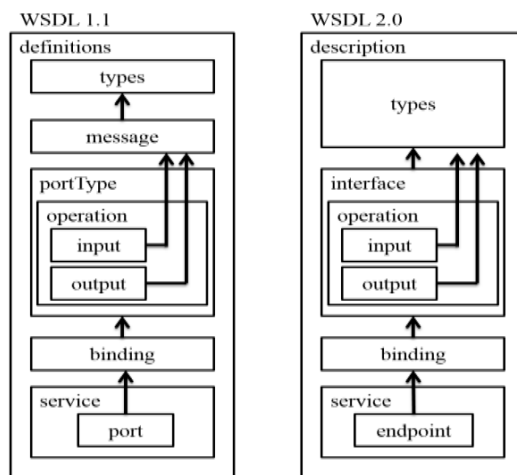


Fig. 1. WSDL 1.1 y WSDL 2.0.

### 3. Trabajos relacionados

Las descripciones de los servicios web, OWLS o WSDL, contienen texto libre (lenguaje natural), además de nombres de parámetros, nombres de datos, complejos y simples, nombres de operaciones y parámetros de entrada-salida. Diversos trabajos han utilizado esta información para una gran variedad de tareas relacionadas al procesamiento de textos. Una de estas tareas es calcular la similitud semántica entre servicios [2, 3, 8, 10, 15], clasificar los servicios basada en su contenido [4, 5, 11, 12, 14, 17, 19] y agrupamiento de servicios Web [16].

En el contexto de la clasificación de servicios Web, se han propuesto enfoques utilizando la colección de servicios OWLS-TC para una clasificación supervisada. El trabajo presentado en [19] utiliza las descripciones del servicio web, las descripciones de las operaciones y las descripciones de las entradas/salidas, con la finalidad de clasificar los servicios; en [14] se utiliza el algoritmo de máquinas de soporte vectorial, características de similitud entre los términos utilizando *WordNet* y una taxonomía del dominio, para clasificar, de manera eficiente, los servicios web de la colección; y en [9] se ha propuesto la clasificación automática de servicios web basada en la teoría de conjuntos aproximados, ellos proponen una representación de las descripciones con espacios vectoriales y una ponderación de los términos basada en la entropía.

No sólo existen trabajos que utilicen la colección mencionada, también, se han propuesto trabajos bajo colecciones privadas de servicios Web. En [11] utilizan una colección de 364 servicios web, obtenidos de repositorios públicos en la Web, para evaluar su metodología propuesta en la clasificación y predicción de la calidad de un servicio web basada en un conjunto de atributos de calidad, a saber: fiabilidad, documentación, rendimiento, respuesta y tiempo. En el contexto de la clasificación no supervisada, el trabajo presentado en [17] propone método para la clasificación automática a partir de las etiquetas presentes en el documento de descripción de cada

servicio, su método fue probado con 951 archivos WSDL distribuidos en 19 categorías. El trabajo propuesto en [12] presenta un enfoque de minería de textos para la clasificación automática de servicios web en dominios específicos y la identificación de conceptos clave en la documentación textual de los servicios. Su enfoque fue validado con un conjunto de 600 servicios web categorizados en 8 clases. Finalmente, el trabajo expuesto en [4] propone un mecanismo basado en reglas heurísticas que permite a los autores de servicios Web, clasificar sus servicios en una taxonomía manejada por un repositorio.

#### 4. Enfoque propuesto

Este artículo presenta la clasificación de servicios Web utilizando sus descripciones textuales y nombres de parámetros. Esta problemática se aborda mediante el paradigma de clasificación de textos, tarea que consiste en asociar categorías predefinidas a partir del análisis de un texto. En la Fig. 2 se muestra la arquitectura general del enfoque propuesto.

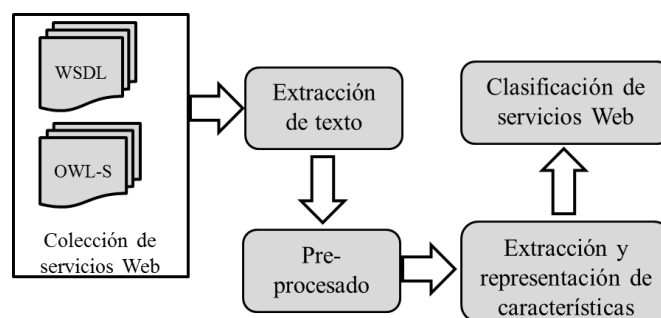


Fig. 2. Arquitectura para la clasificación de Servicios Web.

El enfoque analiza una colección de servicios Web para su clasificación. La colección se conforma de la descripción en WSDL y de su extensión semántica en OWL-S de cada servicio. Bajo esta arquitectura, la primera etapa es la extracción de textos a partir de los archivos de descripción de los servicios Web, para ello, se realiza un analizador de archivos WSDL y OWL-S con la finalidad de extraer el texto en lenguaje natural y los nombres de los elementos relevantes. Después, los textos extraídos son pre-procesados para obtener las unidades léxicas (palabras simples) que describen a cada servicio. El conjunto de unidades léxicas, características de cada servicio web, son representadas como *1-gramas*, *2-gramas* o *3-gramas*, mediante una ponderación de los términos, en un modelo espacio vectorial [1]. Finalmente, los vectores son utilizados para la clasificación de los servicios mediante una red neuronal de tipo perceptrón multicapa.

##### 4.1. Extracción de textos

Los documentos de descripción (WSDL) y los documentos de descripción semántica de la funcionalidad (OWL-S), a partir de la colección de servicios Web,

son analizados con la finalidad de identificar y extraer la información textual, la cual es de utilidad para lograr la clasificación basada en el contenido. De esta manera, se extrae el contenido de las etiquetas *serviceName* y *textDescription* de la clase *Profile* a partir de las descripciones de la funcionalidad (OWL-S), estas etiquetas contienen texto en lenguaje natural. Por su parte, a partir del WSDL se extrae el nombre del servicio (*wSDL:service name*), el nombre de las operaciones (*wSDL:operation name*) y los nombres de los tipos de datos de los mensajes, ya sea simples o complejos (*xsd:simpleType name* y *xsd:complexType name*). Esta información en lenguaje natural y los nombres de los elementos del servicio se utilizan para representar y, posteriormente, clasificar los servicios basándose en su contenido.

## 4.2. Preprocesamiento de los textos

La primera tarea, después de tener el texto de los servicios, es la segmentación en palabras. En el caso de los servicios web, es común encontrar nombres de servicios, operaciones o tipos de datos con palabras compuestas. Para lograr la segmentación de las palabras, se transforman los nombres compuestos en su forma canónica simple, es decir, en unidades léxicas, considerando el cambio de minúscula a mayúscula y el subguión como separadores de palabras. Por ejemplo, *getAddressLocation* o *get\_address\_location* se descomponen en las siguientes unidades léxicas: [*get*] [*Address*] [*Location*].

Adicionalmente, los textos de los servicios se normalizan, mediante la aplicación de una conversión a minúsculas, eliminación de signos de puntuación y supresión de las palabras vacías, es decir, palabras que no aportan significado y por lo tanto se consideran no funcionales para la clasificación de servicios basada en el contenido.

## 4.3. Extracción y representación de las características de los servicios

El conjunto de unidades léxicas (palabras) normalizadas y filtradas son representadas en el modelo espacio vectorial [1]. Este modelo se utiliza para representar los textos de una manera formal utilizando los términos como características, las cuales pueden ser unidades léxicas simples (*1-gramas*), secuencias de dos palabras (*2-gramas*) o cualquier secuencia de palabras (*n-gramas*).

En este artículo complementamos el modelo de espacio vectorial con la representación de bolsa de palabra, por sus siglas en inglés BOW [13], el cual consiste en una colección de textos y su vocabulario de términos (características). Cada texto de los servicios web es representado como un vector  $S_j = (w_{1j}, w_{2j} \dots w_{nj})$ , donde cada componente  $w_{ij}$  expresa la importancia o frecuencia ponderada que produce la característica  $i$ , unidad léxica (*1-grama*) o secuencia de palabras (*n-grama*), del vocabulario en el texto  $j$  del servicio Web. En este trabajo nos centramos en medir la influencia de *1-gramas*, *2-gramas* y *3-gramas* de palabras como características (términos del vocabulario) dentro del espacio vectorial para la clasificación de servicios Web.

Existen diferentes enfoques para obtener la importancia o ponderación de los términos del vocabulario sobre un texto. Este artículo se basa en un pesado *booleano*,

un pesado de frecuencia de aparición del término ( $TF$ ) y un pesado basado en la frecuencia de aparición del término en la colección de textos ( $TF-IDF$ ).

La ponderación *booleana* de un término calcula el peso asignando un valor de 0 si el término del vocabulario no aparece en el texto y un valor de 1 si el término está presente. Mientras que, la ponderación basada en la frecuencia de aparición ( $TF$ ) calcula el número de veces que un término del vocabulario aparece en un texto:  $w_{ij} = TF(t_i, S_j)$ . Esta ponderación otorga importancia a los términos más frecuentes, sin embargo no es una ponderación normalizada lo que puede provocar valores alejados entre los componentes del vector.

Finalmente, también se tiene la ponderación basada en la frecuencia de aparición del término en la colección de textos ( $TF-IDF$ ), la cual captura la importancia que tiene un término para un texto de descripción de servicio Web. Esta ponderación utiliza la frecuencia de aparición de un término del vocabulario en un texto  $TF(t_i, S_j)$  y la frecuencia inversa que determina si el término es común en la colección de textos  $IDF(t_i, S_j) = \log \frac{|S|}{1 + |S \in S : t_i \in S|}$ . Entonces, la fórmula final para el cálculo de  $TF-IDF$  queda de la siguiente manera:  $w_{ij} = TF(t_i, S_j) \times IDF(t_i, S_j)$ .

#### 4.4. Clasificación de servicios web

La clasificación de servicios web se basa en los vectores de cada servicio ponderado de manera *booleana*,  $TF$  o  $TF-IDF$ . La clasificación presentada en este artículo se considera una clasificación supervisada, ésta ha sido ampliamente utilizada en el campo del aprendizaje automático para estimar la función predictora de cada clase de nuestra colección. Por ello, se dividirá la colección de servicios web en un conjunto de entrenamiento y un conjunto para las pruebas.

El objetivo de esta etapa es construir un clasificador de servicios web considerando 9 categorías: *Communication* (Comunicación), *Economy* (Economía), *Education* (Educación), *Food* (Comida), *Geography* (Geografía), *Medical* (Medicina), *Simulation* (Simulación), *Travel* (Turismo) y *Weapon* (Armamento).

La tarea de clasificación se lleva a cabo mediante una red neuronal artificial de tipo perceptrón multicapa con aprendizaje supervisado. El perceptrón multicapa es una Red Neuronal Artificial (RNA) formada por múltiples capas de neuronas cuya finalidad es resolver problemas de múltiples clases. Este tipo de redes neuronales, presentadas en [7], [9] y [18], son considerados algoritmos de clasificación supervisada y están inspirados en las redes neuronales biológicas del cerebro humano. Bajo un esquema matemático, un perceptrón multicapa es una función no lineal compleja con un conjunto de unidades neuronales interconectadas, compuesta de una capa de entrada y una de salida, esta red es entrenada con un conjunto de servicios y, entonces, realiza los ajustes de parámetros correspondientes para arrojar salidas similares con los datos de prueba.

La idea es evaluar la tarea de clasificación de servicios con el perceptrón multicapa en combinación con las ponderaciones (*booleana*,  $TF$  o  $TF-IDF$ ) y la formación de términos (*1-gramas*, *2-gramas* y *3-gramas*), con la finalidad de encontrar la mejor solución en términos de precisión. La implementación de la red neuronal artificial de tipo perceptrón multicapa con aprendizaje supervisado, para la clasificación de servicios Web, se ha llevado a cabo mediante la herramienta WEKA [6].

## 5. Experimentación y resultados

La evaluación del enfoque propuesto fue realizada con la versión 3.0 de la colección OLWS-TC<sup>7</sup>, la cual consta de 1129 servicios descritos usando WSDL y OWL-S. Los servicios de esta colección está preclasificados en las siguientes categorías: *Communication* (Comunicación), *Economy* (Economía), *Education* (Educación), *Food* (Comida), *Geography* (Geografía), *Medical* (Medicina), *Simulation* (Simulación), *Travel* (Turismo) y *Weapon* (Armamento). Esta colección fue dividida en dos grupos: 899 para el aprendizaje del modelo de clasificación y 230 para las pruebas.

La experimentación consiste en evaluar el algoritmo de clasificación, red neuronal artificial de tipo perceptrón multicapa, combinándolo con la representación basada en *1-gramas*, *2-gramas* y *3-gramas*, además de evaluar los tres tipos de ponderaciones para los términos: *booleano*, *TF* y *TF-IDF*. Todas las configuraciones de los experimentos fueron ejecutadas sobre el conjunto de 899 descripciones de servicios para lograr el aprendizaje del modelo de predicción y, posteriormente, evaluadas con el conjunto 230 descripciones de prueba. Además, todos los experimentos se realizaron con los siguientes parámetros de la red neuronal de tipo perceptrón multicapa: *número de capas ocultas* =  $\frac{\# \text{ atributos} + \# \text{ clases}}{2}$ ; *tasa de aprendizaje* = 0.3; y, *momentum*=0.2.

La evaluación de todos los experimentos se realizó utilizando las métricas de *Precisión (P)*, *Recuerdo (R)* y *medida F* ampliamente utilizadas en cualquier tarea de clasificación, en nuestro caso, clasificación de textos. Estas métricas comparan los resultados del clasificador a evaluar con los valores externos de confianza (servicios web preclasificados), utilizando los siguientes valores: a) *Verdadero Positivo (VP)* es el número de predicciones correctas del clasificador de servicios que corresponden al juicio externo de confianza (servicios preclasificados); *Verdadero Negativo (VN)* es el número de predicciones correctas del clasificador de servicios que no corresponden al juicio externo de confianza; *Falso Positivo (FP)* corresponde al número predicciones incorrectas del clasificador de servicios que corresponden al juicio externo de confianza; y, finalmente *Falso Negativo (FN)* es el número de predicciones incorrectas del clasificador de servicios que no corresponden al juicio externo de confianza.

Bajo estos criterios, se emplea la *Precisión (P)* para evaluar los algoritmos en términos de los valores de predicciones positivas, la cual se define como:

$$P = \frac{VP}{VP + FP} \quad (1)$$

También, se utiliza el *Recuerdo (R)* para expresar la tasa de correspondencias correctas con los servicios preclasificados de manera externa con una alta confianza, el cual se define como:

$$R = \frac{VP}{VP + FN} \quad (2)$$

<sup>7</sup> <http://projects.semwebcentral.org/projects/owls-tc/>

Finalmente, la *medida F* que representa la media armónica entre *Precisión* y *Recuerdo*, la cual tiene como fundamento el obtener un valor único ponderado entre ellas y se define como:

$$medida F = 2 * \frac{P * R}{P + R} \tag{3}$$

Los experimentos se han organizado para medir el impacto de los tipos de representación de los términos y las ponderaciones utilizadas en las características de las descripciones de los servicios Web. Bajo esta consideración, la Tabla 1 muestra los resultados, por categoría, de los experimentos utilizando la ponderación *booleana* para las características *1-gramas*, *2-gramas* y *3-gramas*. También, se proporciona un peso promedio, resultado de considerar la *medida F* y un factor que refleja la importancia (número de servicios) de cada clase.

**Tabla 1.** Resultados de la clasificación considerando la ponderación *booleana*.

Categoría	1-grama			2-grama			3-grama		
	P	R	F	P	R	F	P	R	F
Comunicación	0.98	0.96	0.97	0.96	0.81	0.88	0.87	0.82	0.84
Economía	0.98	0.97	0.97	0.98	0.97	0.97	0.96	0.95	0.95
Educación	0.90	0.97	0.94	0.91	0.96	0.93	0.89	0.94	0.91
Comida	0.34	0.30	0.32	0.26	0.24	0.25	0.25	0.23	0.24
Geografía	1.00	0.98	0.99	1.00	0.98	0.99	0.98	0.96	0.97
Medicina	0.36	0.40	0.38	0.30	0.33	0.31	0.3	0.31	0.30
Simulación	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.98	0.97
Turismo	0.99	0.92	0.95	0.99	0.96	0.97	0.97	0.94	0.95
Armamento	1.00	0.93	0.96	0.97	0.95	0.96	0.96	0.95	0.95
<b>Peso promedio</b>	0.881	0.881	<b>0.880</b>	0.873	0.869	<b>0.870</b>	0.851	0.852	<b>0.851</b>

En la Tabla 2 se muestran los resultados de la ponderación *TF* para los términos *1-gramas*, *2-gramas* y *3-gramas* para la clasificación de servicios con la red neuronal perceptrón multicapa.

**Tabla 2.** Resultados por categoría utilizando la ponderación *TF*.

Categoría	1-grama			2-grama			3-grama		
	P	R	F	P	R	F	P	R	F
Comunicación	0.92	0.88	0.90	0.90	0.85	0.87	0.85	0.81	0.83
Economía	0.97	0.95	0.96	0.95	0.93	0.94	0.85	0.84	0.84
Educación	0.86	0.86	0.86	0.84	0.84	0.84	0.81	0.82	0.81
Comida	0.24	0.20	0.21	0.22	0.20	0.21	0.20	0.20	0.20
Geografía	0.91	0.85	0.88	0.88	0.82	0.85	0.83	0.78	0.80
Medicina	0.28	0.28	0.28	0.26	0.26	0.26	0.20	0.21	0.20
Simulación	0.85	0.69	0.76	0.82	0.67	0.74	0.76	0.61	0.68
Turismo	0.81	0.95	0.87	0.78	0.90	0.84	0.73	0.81	0.77
Armamento	0.86	0.78	0.82	0.80	0.72	0.76	0.74	0.65	0.69



Categoría	1-grama			2-grama			3-grama		
	P	R	F	P	R	F	P	R	F
<b>Peso promedio</b>	0.815	0.816	<b>0.815</b>	0.792	0.791	<b>0.790</b>	0.732	0.733	<b>0.732</b>

Finalmente, la Tabla 3 presenta los resultados de la ponderación *TF-IDF* para los términos *1-gramas*, *2-gramas* y *3-gramas* en la clasificación de servicios con la red neuronal perceptrón multicapa.

**Tabla 3.** Resultados utilizando la ponderación *TF-IDF*.

Categoría	1-grama			2-grama			3-grama		
	P	R	F	P	R	F	P	R	F
Comunicación	0.98	0.97	0.97	0.97	0.96	0.96	0.92	0.91	0.91
Economía	0.98	0.98	0.98	0.97	0.98	0.97	0.91	0.92	0.91
Educación	0.91	0.97	0.94	0.91	0.97	0.94	0.85	0.89	0.87
Comida	0.59	0.55	0.57	0.41	0.35	0.38	0.31	0.25	0.28
Geografía	0.99	0.98	0.92	0.97	0.97	0.97	0.87	0.88	0.87
Medicina	0.56	0.57	0.56	0.39	0.41	0.40	0.29	0.39	0.33
Simulación	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.96	0.96
Turismo	0.94	0.93	0.93	0.99	0.92	0.95	0.89	0.85	0.87
Armamento	0.98	0.98	0.98	0.99	0.95	0.97	0.93	0.87	0.90
<b>Peso promedio</b>	0.904	0.915	<b>0.910</b>	0.885	0.888	<b>0.886</b>	0.813	0.820	<b>0.816</b>

Los resultados mostrados en las Tablas 1, 2 y 3 hacen notar que la mejor alternativa es considerar de formación de términos *1-gramas* y su ponderación *TF-IDF* para la clasificación de servicios Web mediante una red neuronal perceptrón multicapa. Con esta configuración se logra un 91 % de servicios clasificados correctamente.

Los resultados de la experimentación demuestran la efectividad de nuestro enfoque para la clasificación de servicios Web. A pesar de que los resultados no son alentadores para dos clases (Medicina y Comida), el enfoque puede ayudar a los desarrolladores de servicios web a localizar y reutilizar los componentes de software existentes en un repositorio desorganizado de servicios.

## 6. Conclusiones

Este artículo ha presentado un enfoque de clasificación de servicios web utilizando una red neuronal artificial de tipo perceptrón multicapa y mediante n-gramas de palabras. Se han presentado diversos experimentos, en los cuales se combinaron la composición de términos del vocabulario (*1-gramas*, *2-gramas* y *3-gramas*) con las ponderaciones de estos términos (*booleano*, frecuencia de aparición, y frecuencia de aparición inversamente proporcional a su aparición en toda la colección). A partir de esta experimentación, se hace notar que nuestra propuesta de clasificación de servicios web ha logrado un 91 % de eficiencia.

Las principales contribuciones de este trabajo son: a) las combinaciones de composiciones de términos y sus ponderaciones en la experimentación para la clasificación de servicios Web; b) la extracción de términos de la colección para formar el vocabulario; y c) el modelo de clasificación, basado en una red neuronal perceptrón multicapa para la clasificación de servicios Web mediante sus descripciones OWLS y WSDL.

Como trabajo futuro, la clase descubierta por el clasificador puede ser almacenada, mediante la instanciación en una ontología para tareas de extracción de información, descubrimiento y composición de servicios Web.

## Referencias

1. Baeza-Yates, R., Ribeiro-Neto, B.: *Modern Information Retrieval*. Addison Wesley, Boston (1999)
2. Bravo, M., Alvarado, M.: Similarity measures for substituting Web services. *International Journal of Web Services Research*, 7 (3), pp. 1–29 (2010)
3. Bravo, M., Rodríguez, J., Reyes, A.: Enriching Semantically Web Service Descriptions. In: *On the Move to Meaningful Internet Systems: OTM 2014 Conferences*, pp. 776–783, Springer Berlin Heidelberg (2014)
4. Corella, M.Á., Castells, P.: A heuristic approach to semantic web services classification. In: *Knowledge-Based Intelligent Information and Engineering Systems*, pp. 598–605 (2006)
5. Chen, L., Zhang, Y., Song, Z.L., Miao, Z.: Automatic web services classification based on rough set theory. *Journal of Central South University*, 20, pp. 2708–2714 (2013)
6. Garner, S.R.: Weka: The Waikato environment for knowledge analysis. In: *Proc. of the New Zealand Computer Science Research Students Conference*. pp. 57–64 (1995)
7. Hassoun, M.H.: *Fundamentals of Artificial Neural Networks*. Cambridge, MA: MIT press (1999)
8. Hau, J., Lee, W., Darlington, J.: A semantic similarity measure for semantic web services. In: *Web Service Semantics Workshop at WWW*, pp. 10–14 (2005)
9. Haykin, S.: *Neural Networks and Learning Machines*. 3rd Edition, Pearson Education, New Jersey (2008)
10. Liu, F., Shi, Y., Yu, J., Wang, T., Wu, J.: Measuring similarity of web services based on WSDL. In: *2010 IEEE International Conference on Web Services*, pp. 155–162 (2010)
11. Mohanty, R., Ravi, V., Patra, M.R.: Web-services classification using intelligent techniques. *Expert Systems with Applications*, 37(7), pp. 5484–5490 (2010)
12. Nisa, R., Qamar, U.: A text mining based approach for web service classification. *Information Systems and e-Business Management*, pp. 1–18 (2014)
13. Sebastiani, F.: Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), pp. 1–47 (2002)
14. Wang, H., Shi, Y., Zhou, X., Zhou, Q., Shao, S., Bouguettaya, A.: Web service classification using support vector machine. In: *22nd IEEE International Conference on Tools with Artificial Intelligence*, Vol. 1, pp. 3–6 (2010)
15. Wu, J., Wu, Z.: Similarity-based web service matchmaking. In: *IEEE International Conference on Computing*, Vol. 1, pp. 287–294 (2005)
16. Wu, J., Chen, L., Zheng, Z., Lyu, M. R., Wu, Z.: Clustering web services to facilitate service discovery. *Knowledge and information systems*, 38(1), pp. 207–229 (2014)

17. Yuan-jie, L., Jian, C.: Web service classification based on automatic semantic annotation and ensemble learning. In: 26th International on Parallel and Distributed Processing Symposium Workshops & PhD Forum, pp. 2274–2279 (2012)
18. Zak, S.H.: Systems and Control. University Press, New York (2003)
19. Zhang, J., Pan, D.: Web Service Classification. (2008)